

PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2002-304401**

(43)Date of publication of application : **18.10.2002**

(51)Int.Cl.

G06F 17/30

G06F 17/60

(21)Application number : **2001-107197**

(71)Applicant : **TOSHIBA CORP**

(22)Date of filing : **05.04.2001**

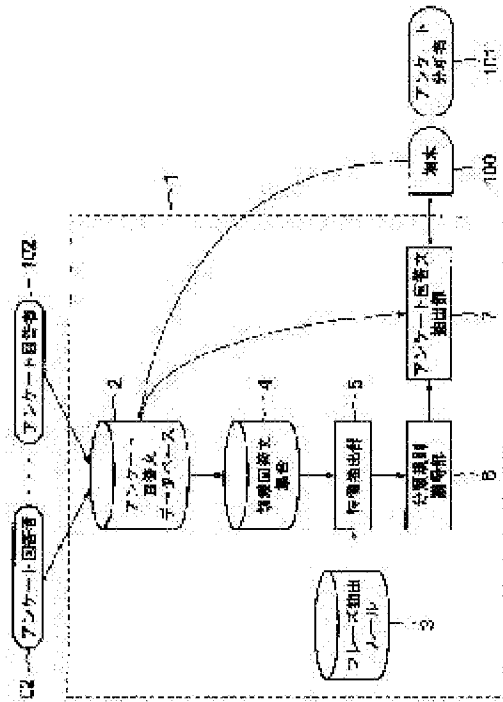
(72)Inventor : **SUYAMA AKIHIRO
SAKURAI SHIGEAKI
ICHIMURA YUMI
ORIHARA RYOHEI**

(54) DEVICE AND METHOD FOR PROCESSING QUESTIONNAIRE AND PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a device for processing questionnaire to classify questionnaire into optional categories in a manner not to manually register complaint texts into a dictionary.

SOLUTION: Training answer text groups 4 containing segmented free questionnaire answer texts to be used for training examples are extracted from a questionnaire answer test database 2. Questionnaire analyzers add classified categories such as complaint, demand or impression to each segment. At a characteristic extracting section 5, the groups 4 and phrase extracting rules 3 are input and characteristic vector groups are formed. At a classification rule acquiring section 6, on the basis of the training examples consisting of plural characteristic vectors and the classified categories, classification rules for providing the classified categories from the characteristic vectors are learned. Using the acquired classification rules, at a questionnaire answer text extracting section 7, the free questionnaire answer texts are classified. The free questionnaire answer texts belonging to the classified categories indicated by the questionnaire analyzers are extracted and presented.



LEGAL STATUS

[Date of request for examination] 06.02.2003

[Date of sending the examiner's decision of rejection] 28.03.2006

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2002-304401

(P2002-304401A)

(43)公開日 平成14年10月18日(2002. 10. 18)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード [*] (参考)
G 0 6 F 17/30	2 1 0 1 7 0	G 0 6 F 17/30	2 1 0 D 5 B 0 7 5 1 7 0 A 1 7 0 Z 2 1 0 A
17/60	2 1 0 1 5 2	17/60	1 5 2
審査請求 未請求 請求項の数9 O L (全 11 頁)			

(21)出願番号 特願2001-107197(P2001-107197)

(22)出願日 平成13年4月5日(2001.4.5)

(71)出願人 000003078

株式会社東芝

東京都港区芝浦一丁目1番1号

(72)発明者 酢山 明弘

神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

(72)発明者 櫻井 茂明

神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

(74)代理人 100058479

弁理士 鈴江 武彦 (外6名)

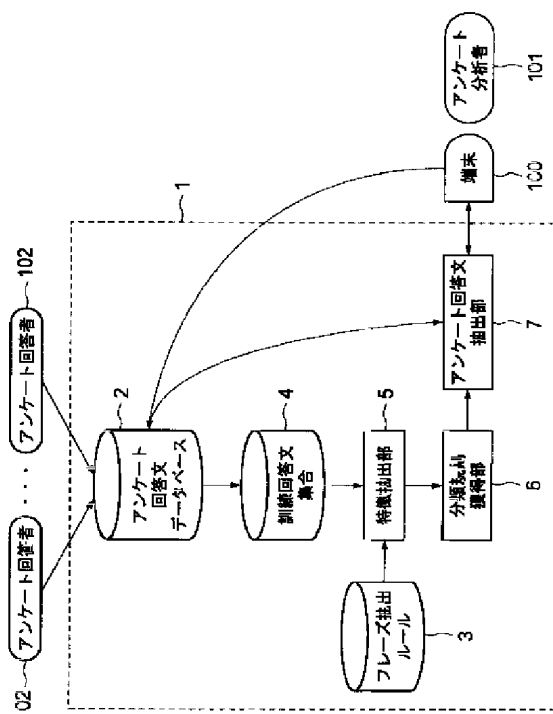
最終頁に続く

(54)【発明の名称】 アンケート処理装置及びアンケート処理方法並びにプログラム

(57)【要約】

【課題】 手作業による苦情文の辞書登録を行わない方法で、アンケートを任意のカテゴリに分類するアンケート処理装置を提供すること。

【解決手段】 アンケート回答文データベース2から訓練例に用いるセグメント化されたフリーアンケート回答文を含む訓練回答文集合4を抽出する。各セグメントには、アンケート分析者が、苦情、要望、感想などの分類カテゴリを付与する。特徴抽出部5では、訓練回答文集合4とフレーズ抽出ルール3を入力として、特徴ベクタ集合を作成する。分類規則獲得部6では、複数の特徴ベクタと分類カテゴリからなる訓練例をもとに、特徴ベクタから分類カテゴリを求めるための分類規則を学習する。そして、獲得された分類規則を用いて、アンケート回答文抽出部7により、フリーアンケート回答文を分類する。また、アンケート分析者が指定した分類カテゴリに属するフリーアンケート回答文を抽出して提示する。



【特許請求の範囲】

【請求項1】フリーアンケート回答文をそれに含まれる内容に応じて分類するためのアンケート処理装置であって、
分類規則生成のために用意された複数のフリーアンケート回答文をもとにして、カテゴリ分類の基準となる表現の候補を複数種類抽出するための抽出手段と、
前記複数のフリーアンケート回答文の各々について、当該フリーアンケート回答文に前記抽出手段により抽出された個々の前記表現の候補に相当する内容が含まれるか否かを示す特徴情報に変換するための変換手段と、
前記複数のフリーアンケート回答文の各々に対応する前記特徴情報及び該フリーアンケート回答文の各々に対して指定された分類カテゴリをもとにして、分類対象のフリーアンケート回答文に該当する分類カテゴリを求めるための分類規則を生成するための生成手段とを備えたことを特徴とするアンケート処理装置。

【請求項2】前記抽出手段は、前記カテゴリ分類の基準となる表現の候補として、名詞句表現又は文末語尾表現に相当するものを抽出することを特徴とする請求項1に記載のアンケート処理装置。

【請求項3】前記特徴情報は、前記表現の候補の種類数と同数のビット列であり、
前記ビット列の個々のビットは、当該ビットに対応する前記表現の候補に相当する内容が、対応する前記フリーアンケート回答文に含まれるか否かを示すものであることを特徴とする請求項1に記載のアンケート処理装置。

【請求項4】前記生成手段は、前記特徴情報及び前記分類カテゴリ並びに前記表現の候補を入力とした帰納学習によって前記分類規則を導くものであることを特徴とする請求項1に記載のアンケート処理装置。

【請求項5】前記分類規則は、前記表現に基づく決定木により前記分類カテゴリを与えるものであることを特徴とする請求項1に記載のアンケート処理装置。

【請求項6】フリーアンケート回答文を含むアンケート回答文を複数記憶するための第1の記憶手段と、
前記分類規則に基づいて、前記記憶手段に記憶された分類対象のフリーアンケート回答文を複数の分類カテゴリのいずれかに分類するための分類手段と、
前記分類手段により分類されたフリーアンケート回答文と、該フリーアンケート回答文に該当する分類カテゴリと、該フリーアンケート回答文を含む前記アンケート回答文を識別する識別情報とを対応付けて記憶するための第2の記憶手段とを更に備えたことを特徴とする請求項1に記載のアンケート処理装置。

【請求項7】所望の分類カテゴリが指定された抽出要求を受け付けるための手段と、
この手段により前記抽出要求が受け付けられた場合に、前記第2の記憶手段を参照して、指定された前記分類カテゴリに該当するフリーアンケート回答文を抽出する手

段と、

この手段により抽出された前記フリーアンケート回答文に関する所定の表示を行うための手段とを更に備えたことを特徴とする請求項6に記載のアンケート処理装置。

【請求項8】フリーアンケート回答文をそれに含まれる内容に応じて分類するためのアンケート処理装置におけるアンケート処理方法であって、
分類規則生成のために用意された複数のフリーアンケート回答文をもとにして、カテゴリ分類の基準となる表現の候補を複数種類抽出し、
前記複数のフリーアンケート回答文の各々について、当該フリーアンケート回答文に前記抽出手段により抽出された個々の前記表現の候補に相当する内容が含まれるか否かを示す特徴情報に変換し、
前記複数のフリーアンケート回答文の各々に対応する前記特徴情報及び該フリーアンケート回答文の各々に対して指定された分類カテゴリをもとにして、分類対象のフリーアンケート回答文に該当する分類カテゴリを求めるための分類規則を生成することを特徴とするアンケート処理方法。

【請求項9】フリーアンケート回答文をそれに含まれる内容に応じて分類するためのアンケート処理装置としてコンピュータを機能させるためのプログラムであって、
分類規則生成のために用意された複数のフリーアンケート回答文をもとにして、カテゴリ分類の基準となる表現の候補を複数種類抽出するための抽出機能と、
前記複数のフリーアンケート回答文の各々について、当該フリーアンケート回答文に前記抽出機能により抽出された個々の前記表現の候補に相当する内容が含まれるか否かを示す特徴情報に変換するための変換機能と、
前記複数のフリーアンケート回答文の各々に対応する前記特徴情報及び該フリーアンケート回答文の各々に対して指定された分類カテゴリをもとにして、分類対象のフリーアンケート回答文に該当する分類カテゴリを求めるための分類規則を生成するための生成機能とをコンピュータに実現させるためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、利用者が指定した任意のカテゴリにフリーアンケート回答文を分類するアンケート処理装置及びアンケート処理方法に関する。

【0002】

【従来の技術】アンケート文から任意のカテゴリに属する文を抽出する技術としては、例えば、特開2000-181926号公報（第2頁～第17頁参照）に記載されている方法が知られている。

【0003】

この方法では、予め人の手によって苦情文を辞書（苦情文辞書）に登録しておく。そして、電子メールが入力されると、上記苦情文辞書を参照して、入力電子メールから、苦情文を含む電子メールを抽出する。

苦情文を含む電子メールが抽出されたならば、その苦情文をその電子メールのID（識別情報）と関連付けて記憶する。苦情文については種別が分けられており、記憶した苦情文を含む電子メールを表示する際に、苦情文の部分を強調表示あるいは種別毎に区別して強調表示している。これによって、入力電子メールから苦情の電子メールを抽出すると共にその苦情内容を種別に分けて検索・集計および苦情内容の指摘を行うことができる。

【0004】

【発明が解決しようとする課題】しかしながら、従来の苦情文抽出処理においては、予め利用者が手作業により苦情文を辞書に登録しなければならないという問題点があった。このため、苦情文の辞書登録に多大なる労力を要するとともに、必ずしも有効な苦情文辞書を作成することができず、抽出能力や分類能力にも問題があった。

【0005】本発明は、上記事情を考慮してなされたもので、手作業による苦情文の辞書登録を行わない方法で、アンケートを任意のカテゴリに分類するアンケート処理装置及びアンケート処理方法を提供することを目的とする。

【0006】

【課題を解決するための手段】本発明は、フリーアンケート回答文をそれに含まれる内容に応じて分類するためのアンケート処理装置であって、分類規則生成のために用意された複数のフリーアンケート回答文をもとにして、カテゴリ分類の基準となる表現の候補を複数種類抽出するための抽出手段と、前記複数のフリーアンケート回答文の各々について、当該フリーアンケート回答文に前記抽出手段により抽出された個々の前記表現の候補に相当する内容が含まれるか否かを示す特徴情報に変換するための変換手段と、前記複数のフリーアンケート回答文の各々に対応する前記特徴情報及び該フリーアンケート回答文の各々に対して指定された分類カテゴリをもとにして、分類対象のフリーアンケート回答文に該当する分類カテゴリを求めるための分類規則を生成するための生成手段とを備えたことを特徴とする。

【0007】好ましくは、前記抽出手段は、前記カテゴリ分類の基準となる表現の候補として、名詞句表現又は文末語尾表現に相当するものを抽出するようにしてもよい。

【0008】好ましくは、前記生成手段は、前記特徴情報及び前記分類カテゴリ並びに前記表現の候補を入力とした帰納学習によって前記分類規則を導くものであるようにしてもよい。

【0009】好ましくは、フリーアンケート回答文を含むアンケート回答文を複数記憶するための第1の記憶手段と、前記分類規則に基づいて、前記記憶手段に記憶された分類対象のフリーアンケート回答文を複数の分類カテゴリのいずれかに分類するための分類手段と、前記分類手段により分類されたフリーアンケート回答文と、該

フリーアンケート回答文に該当する分類カテゴリと、該フリーアンケート回答文を含む前記アンケート回答文を識別する識別情報とを対応付けて記憶するための第2の記憶手段とを更に備えるようにしてもよい。

【0010】また、本発明は、フリーアンケート回答文をそれに含まれる内容に応じて分類するためのアンケート処理装置におけるアンケート処理方法であって、分類規則生成のために用意された複数のフリーアンケート回答文をもとにして、カテゴリ分類の基準となる表現の候補を複数種類抽出し、前記複数のフリーアンケート回答文の各々について、当該フリーアンケート回答文に前記抽出手段により抽出された個々の前記表現の候補に相当する内容が含まれるか否かを示す特徴情報に変換し、前記複数のフリーアンケート回答文の各々に対応する前記特徴情報及び該フリーアンケート回答文の各々に対して指定された分類カテゴリをもとにして、分類対象のフリーアンケート回答文に該当する分類カテゴリを求めるための分類規則を生成することを特徴とする。

【0011】なお、装置に係る本発明は方法に係る発明としても成立し、方法に係る本発明は装置に係る発明としても成立する。また、装置または方法に係る本発明は、コンピュータに当該発明に相当する手順を実行させるための（あるいはコンピュータを当該発明に相当する手段として機能させるための、あるいはコンピュータに当該発明に相当する機能を実現させるための）プログラムとしても成立し、該プログラムを記録したコンピュータ読取り可能な記録媒体としても成立する。

【0012】本発明によれば、フリーアンケート回答文のサンプルから、対象となるフリーアンケート回答文を分類することができる分類規則を学習することができるため、手作業による苦情文などの辞書登録を行わない方法で、アンケートを任意のカテゴリに分類することが可能になる。

【0013】

【発明の実施の形態】以下、図面を参照しながら発明の実施の形態を説明する。

【0014】なお、アンケートに対する回答のうち、自由記述による回答文を、「フリーアンケート回答文」と呼ぶものとする。

【0015】図1に、本発明の一実施形態に係るアンケート処理装置の構成例を示す。

【0016】本アンケート処理装置1は、フリーアンケート回答文の内容から分類カテゴリ（例えば、苦情、要望、感想など）を求める分類規則を学習するための機能や、アンケート分析者等のユーザ（以下、単にアンケート分析者と呼ぶ）（101）が抽出しようとしている任意の分類カテゴリに属するフリーアンケート回答文の提示・記録などを行うための機能等を有するものである。

【0017】図1に示されるように、本アンケート処理装置1は、アンケート回答文データベース2、フレーズ

抽出ルール3、訓練回答文集合4、特徴抽出部5、分類規則獲得部6、アンケート回答文抽出部7を備えている。

【0018】なお、ユーザインタフェース手段については、例えば、アンケート処理装置1にLAN等のネットワークを介して接続された端末100にてアンケート分析者等が操作を行う形態（サーバ・クライアント型）であってもよいし、アンケート処理装置1の持つ入出力装置（図示せず）にてアンケート分析者等が操作を行う形態（スタンドアロン型）であってもよいし、両者が可能であってもよい。

【0019】アンケート回答文データベース2には、アンケート回答者（102）から寄せられたアンケート回答文が登録される。

【0020】アンケート回答者からアンケート回答文を受け取る手段は、例えば、インターネットなどのネットワークを介する方法であってもよいし、記録媒体を介する方法であってもよい。また、例えば回答者から電話で寄せられた回答に基づいて本アンケート処理装置1上でアンケート回答文を作成・入力する方法も可能である。もちろん、複数の方法を併用してもよい。

【0021】例えば、インターネットのホームページを利用したWEBアンケートなどの手段により図2に例示するようなアンケートが行われ、これに対してアンケート回答者（102）から寄せられた回答が、図3に例示するように「行」をアンケート回答者とし、「列」を質問とし、「セル」を各回答者の各質問に対する回答として、アンケート回答文データベース2に登録される。

【0022】フレーズ抽出ルール3は、フリーアンケートの分類に必要な情報を抽出する規則を、形態素の正規表現により記述したものである。図4に、フレーズ抽出ルールの一例を示す。図4では、フリーアンケート回答文の「名詞句表現」（ID＝“＃R1”）と「文末表現」（ID＝“＃R1”）を抽出する規則を例示している。

【0023】訓練回答文集合4は、アンケート回答文データベース2に含まれるフリーアンケート回答文に対する部分集合を、以下の情報も加えて記述したものである。・アンケート回答文ID・フリーアンケート回答文・分類カテゴリただし、分類カテゴリは、アンケート分析者（101）により付与される。

【0024】図5に、訓練回答文集合の一例を示す。なお、図5では、アンケート回答文ID（例えば、Q3＃1）は、図3に例示したアンケート回答文データベース2の「列」（例えば、Q3）と「セル」（例えば、＃1）に対応させている。また、＃T1、＃T2、＃T3、…は、各訓練回答文のIDである。

【0025】以下、本実施形態のアンケート処理装置1の動作手順について説明する。

【0026】図6に、アンケート処理装置1の動作手順

の一例を示す。

【0027】本実施形態のアンケート処理装置1の処理は、図6（a）に例示したような学習フェーズと、図6（b）に例示したような運用フェーズとに分けられる。

【0028】学習フェーズは、アンケート処理装置1をはじめて利用する場合や分類規則を修正する場合に行われる処理である。運用フェーズは、アンケート分析者（101）がアンケート分析を行う場合の処理である。

【0029】（学習フェーズ）学習フェーズにおいて、アンケート分析者（101）は、予め分類カテゴリ（例えば、苦情、要望、感想など）を設定する（ステップS1）。例えば、分類カテゴリ設定画面を表示して分類カテゴリの入力を受け付け、アンケート分析者（101）により入力された分類カテゴリを格納する。あるいは、例えば、分類カテゴリ設定画面にて種々の分類カテゴリを表示し、アンケート分析者（101）による選択入力を受け付ける。

【0030】次に、アンケート回答文データベース2からフリーアンケート回答文のみを抽出し、さらに適度な部分集合を抽出する（ステップS2）。適度な部分集合としてどのくらいの量を抽出するかについては、例えば、抽出するサンプル数あるいはサンプル数／全回答文の比率を、その都度、アンケート分析者（101）が指示するか、あるいは、予め設定しておくなどする方法によってもよい。

【0031】次に、フリーアンケート回答文の部分集合の各フリーアンケート回答文を、『。』や『』（後者は空白文字）のような句点により分割し、セグメント化を行う（ステップS3）。セグメント化に関しては他にも逆接語、『。』読点などによる分割等、他の方法も考えられる。

【0032】そして、アンケート分析者（101）は、セグメント化された各々のフリーアンケート回答文に対して、上記の分類カテゴリのうちから適切なものを選択して付加し、訓練回答文集合4を完成させる（ステップS4）。例えば、分類カテゴリ割り当て画面にて、セグメント化されたフリーアンケート回答文および分類カテゴリの一覧を表示し、アンケート分析者（101）により選択入力された（割り当てられた）分類カテゴリを、当該セグメント化されたフリーアンケート回答文に付与することを繰り返すことによって、訓練回答文集合4を完成させる。

【0033】なお、ステップS1～S4までの処理において、図3が入力データ例に相当し、図5が出力データ例に相当する。

【0034】ステップS4で得られた訓練回答文集合4は、後の再学習のために保存しておくのが好ましい。

【0035】続いて、特徴抽出部5では、詳しくは後述するように、図5のような訓練回答文集合4と図4のようなフレーズ抽出ルール3を入力として、分類規則獲得

部6の入力となるフリーアンケート特徴ベクタ集合（図10参照）を作成する（ステップS5）。

【0036】続いて、分類規則獲得部6では、詳しくは後述するように、フリーアンケート特徴ベクタ集合を入力として、分類規則（図13参照）を学習し、記録する（ステップS6）。

【0037】（運用フェーズ）運用フェーズでは、分類規則獲得部6によって獲得された分類規則を用いてアンケート回答文抽出部7により、フリーアンケート回答文を分類し、アンケート分析者（101）が必要とするカテゴリに属するフリーアンケート回答文を抽出することができる（ステップS11）。

【0038】ここで、本実施形態では、再学習を行うこともできるようにしている。

【0039】すなわち、アンケート分析者（101）は、ステップS11で抽出されたフリーアンケート回答文を見て、明らかに誤分類されたフリーアンケート回答文が存在するか確認する（ステップS12）。

【0040】存在しない場合は、処理を終了するが、誤分類されたフリーアンケート回答文が存在する場合は、セグメント単位で正しい分類カテゴリを付加する（ステップS13）。

【0041】正しい分類カテゴリを付加されたセグメントは、アンケート処理装置1により新たに訓練回答文として登録される（ステップS14）。

【0042】そして、更新された訓練回答文集合4によってステップS5、S6を行って、再学習を実行する。

【0043】以上のように、本実施形態によれば、予め分類規則を与えることなく任意のカテゴリに分類、抽出することが可能なアンケート処理装置を実現することができる。

【0044】さて、以下では、特徴抽出部5、分類規則獲得部6、アンケート回答文抽出部7についてそれぞれさらに詳しく説明する。

【0045】まず、特徴抽出部5について説明する。

【0046】図7に、特徴抽出部5の構成例を示す。図7に示されるように、本特徴抽出部5は、形態素解析部11、その出力である形態素つき訓練回答文集合12、キーフレーズ抽出部13、その出力であるキーフレーズベクタ14、フォーマット変換部15、その出力であるフリーアンケート回答文特徴ベクタ集合16を含む。

【0047】形態素つき訓練回答文集合12は、図8に例示するように訓練回答文集合4（図5参照）に形態素のタグがついた形式で与えられる。なお、図8のタグは一例であり、形態素のタグは形態素解析部11の実装に応じて種々の形式のものが使用可能である。

【0048】キーフレーズベクタ14は、図9に例示するように形態素タグのついた表現（キーフレーズ）と、各表現の形態素つき訓練回答文における出現頻度とのペアで記録されている。出現頻度は、当該キーフレーズを

後の処理で使用するか否か破棄するかを決定するために用いるもので、例えば、出現頻度が所定の基準値以上の場合に、当該キーフレーズを後の処理で使用する と決定する。この場合、所定の基準値は、予め定めてもよいし、得られたキーフレーズベクタ14の出現頻度の大きさやばらつきを考慮してその都度決めてもよい（なお、後述する例では、説明を簡単にするために、基準値＝1にしている）。また、例えば、選択するキーフレーズの数を先に決定して、その数に達するまで出現頻度の大きい順にキーフレーズを選択する方法もあるし、その他にも種々の方法が考えられる。

【0049】フリーアンケート回答文特徴ベクタ集合16は、図10に例示するように、下記のフィールドを1レコードに登録したものである。

- ・アンケート回答文ID
- ・特徴ベクタ
- ・分類カテゴリ

特徴ベクタは、キーフレーズベクタ14の各フレーズが当該回答文セグメントに出現した場合の値として1、出現しなかった場合の値として0をとるバイナリ値の列として表現される。例えば、図10では、最上位ビットが図9のID＝#K1のキーフレーズの出現の有無に対応しており、最下位ビットが図9のID＝#K8のキーフレーズの出現の有無に対応している。

【0050】図11に、図7に示した特徴抽出部5の動作手順の一例を示す。

【0051】アンケート回答文データベース2から生成された訓練回答文集合4（図5参照）に対し、形態素解析部11によって形態素解析を行い、形態素つき訓練回答文集合12（図8参照）を得る（ステップS21）。

【0052】次に、キーフレーズ抽出部13では、得られた形態素つき訓練回答文集合12をフレーズ抽出ルール3（図4参照）に照合し、該当する表現をキーフレーズベクタ14（図9参照）に格納する（ステップS22）。

【0053】そして、フォーマット変換部15では、形態素つき訓練回答文集合12の各形態素つき訓練回答文に対して（ステップS23、S24）、上記セグメント化された各形態素つき訓練回答文をフリーアンケート回答文特徴ベクタへと変換し、フリーアンケート回答文特徴ベクタ集合16（図10参照）へと格納する（ステップS25）。

【0054】上記処理手順に従えば、例えば、図5のID＝“#T1”の訓練回答文

“製品Oはよく壊れて困る”

に対しては、ステップS21により図8の

“/製品<名>/O<英>+は<付>/よく形>+く<活尾>/壊れ<下一>+て<付>/困<五ら>+る<活尾>/<句読>”

が得られ、ステップS22により図4のフレーズ抽出ルール#R1および#R2から、それぞれ図9のID＝

“#K1”のキーフレーズ

“/製品<名>/O<英>”

およびID = “#K2”のキーフレーズ

“/困<五ら>+る<活尾>”

を得る。同様の処理を他の訓練回答文に関しても行った結果が図9のキーフレーズベクタである。

【0055】フリーアンケート回答文特徴ベクタの特徴ベクタフィールドは、キーフレーズベクタの先頭から順にそのフレーズが出現した場合を1、しなかった場合を0とした値をベクトル表現として与えたものであるから、上記例の

“/製品<名>/O<英>+は<付>/よ<形>+く<活尾>/壊れ<下→>+て<付>/困<五ら>+る<活尾>/<句読>”

は、ステップS25により図10のID = “#F1”の特徴ベクタのように

“11000000”

と変換される。

【0056】次に、分類規則獲得部6について説明する。

【0057】図12に、分類規則獲得部6の構成例を示す。図9のようなキーフレーズベクタ（特徴集合）14と図10のようなフリーアンケート回答文特徴ベクタ集合（訓練事例）16とを入力として、帰納学習部21により、フリーアンケート回答文を適切なカテゴリに分類するための分類規則データベース22を得る。

【0058】なお、帰納学習部21における帰納学習としては、例えば、特開平5-257694号公報に開示されたファジィ知識獲得装置の手法を用いることができる。もちろん、これに限定されるものではなく、種々の学習手法が利用可能である。

【0059】例えば、図9の例および図10の例を入力して、帰納学習部21により、図13のような決定木による分類規則を得ることができる。図13の分類規則の解釈は、以下のようである。

- ① “/困<五ら>+る<活尾>”が回答文に出現したら“苦情”に分類される。
- ② “/壊れ<下→>+て<付>/困<五ら>+る<活尾>”が出現しないで、かつ、“/よ<形>+い<活尾>”が出現した場合、“感想+”に分類される。
- ③ “/壊れ<下→>+て<付>/困<五ら>+る<活尾>”が出現しないで、かつ、“/よ<形>+い<活尾>”が出現しない場合、“感想-”に分類される。

【0060】以上により、特定のカテゴリを抽出する分類規則をアンケート分析者が予め記述しなくとも、アンケートの分類するための分類規則を獲得することができる。

【0061】次に、アンケート回答文抽出部7について説明する。

【0062】図14に、アンケート分析者（101）が指定した分類カテゴリに属するフリーアンケート回答文

を抽出するアンケート回答文抽出部7の構成例を示す。

【0063】アンケート分析者（101）は、例えば後で説明する図16のような分析インタフェース32を通して、アンケート回答文データベース2のフリーアンケート回答文进行分类することが可能である。推論エンジン31は、その分類処理を行うものである。

【0064】図15に、図14に示したアンケート回答文抽出部7の動作手順の一例を示す。

【0065】ステップS31では、フリーアンケート回答文をアンケート回答文データベース2から抽出する。

【0066】なお、ここでは、すべてのフリーアンケート回答文を抽出する方法と、指定された条件を満たすフリーアンケート回答文を抽出する方法と、それらを選択可能とする方法とがある。条件の指定方法としては、例えば、フリーアンケート回答文に分類カテゴリ以外の属性情報を付与し、アンケート分析者（101）が所望の属性情報を指定する方法がある。

【0067】次に、抽出されたフリーアンケート回答文の集合に対して形態素解析を行う（ステップS32）。ここで用いる形態素解析は、図7の形態素解析部11で利用したものと同一ものとするのが好ましい。

【0068】次に、各々の形態素つきフリーアンケート回答文に対して（ステップS33、S34）、以下の処理を行う。

【0069】すなわち、まず、当該形態素つきフリーアンケート回答文をセグメント化する（ステップS35）。

【0070】次に、当該セグメント化された形態素つきフリーアンケート回答文を、分類規則データベース22の条件部に照合し（ステップS36）、分類規則に照合したならば（ステップS37）、当該セグメント化された形態素つきフリーアンケート回答文と、照合した分類規則の結論部（分類カテゴリ）と、該当するアンケートIDとを対応づけて、データベース33に記録する（ステップS36～S38）。

【0071】上記処理終了後、アンケート分析者（101）が、分析インタフェース32を介して、分析を行う質問の番号や、予め設定した分類カテゴリから分類カテゴリを選択することで、カテゴリに属するアンケート回答文を抽出して分類の判断に至ったセグメントを強調表示して提示したり、誤った結果を修正したり、別記憶領域に保存することが可能となる（ステップS39）。

【0072】図16に、アンケート回答文抽出部7の分析インタフェース32のGUI（グラフィカルユーザインタフェース）画面321のイメージ例を示す。

【0073】本例の分析インタフェース32は、質問文選択ボックス201、分類カテゴリ選択ボックス202、検索ボタン203、抽出回答文表示ウィンドウ204、修正ボタン205、保存ボタン206などのGUI環境をユーザに提供する。

【0074】この場合の実行手順は、以下に示す通りである。

【0075】分析インタフェース32は、質問文選択ボックス201および分類カテゴリ選択ボックス202から分析を行いたいフリーアンケート質問および分類カテゴリを選択する。そして、検索ボタン203を選択することで、抽出回答文表示ウインドウ204に上記設定された条件に照合するフリーアンケート回答文の判断に至ったセグメントを強調表示して出力する。また、結果が明らかに間違っている場合は、修正ボタン205を選択し修正処理を行う。また、保存ボタン205を選択することでファイルに保存される。

【0076】もちろん、上記の例の他にも、種々のインタフェースが実現可能である。

【0077】なお、以上の各機能は、ソフトウェアとして実現可能である。また、本実施形態は、コンピュータに所定の手段を実行させるための（あるいはコンピュータに所定の機能を実現させるための）プログラムとして実施することもでき、該プログラムを記録したコンピュータ読取り可能な記録媒体として実施することもできる。

【0078】なお、この発明の実施の形態で例示した構成は一例であって、それ以外の構成を排除する趣旨のものではなく、例示した構成の一部を他のもので置き換えたり、例示した構成の一部を省いたり、例示した構成に別の機能あるいは要素を付加したり、それらを組み合わせたりすることなどによって得られる別の構成も可能である。また、例示した構成と論理的に等価な別の構成、例示した構成と論理的に等価な部分を含む別の構成、例示した構成の要部と論理的に等価な別の構成なども可能である。また、例示した構成と同一もしくは類似の目的を達成する別の構成、例示した構成と同一もしくは類似の効果を奏する別の構成なども可能である。また、この発明の実施の形態で例示した各種構成部分についての各種バリエーションは、適宜組み合わせて実施することが可能である。また、この発明の実施の形態は、個別装置としての発明、関連を持つ2以上の装置についての発明、システム全体としての発明、個別装置内部の構成部分についての発明、またはそれらに対応する方法の発明等、種々の観点、段階、概念またはカテゴリに係る発明を包含・内在するものである。従って、この発明の実施の形態に開示した内容からは、例示した構成に限定されることなく発明を抽出することができるものである。

【0079】本発明は、上述した実施の形態に限定されるものではなく、その技術的範囲において種々変形して実施することができる。

【0080】

【発明の効果】本発明によれば、手作業による苦情文の辞書登録を行わない方法で、アンケートを任意のカテゴリ

りに分類することができる。

【図面の簡単な説明】

【図1】本発明の一実施形態に係るアンケート処理装置の構成例を示す図

【図2】アンケート質問の一例を示す図

【図3】アンケート回答文データベースの一例を示す図

【図4】フレーズ抽出ルールの一例を示す図

【図5】訓練回答文集合の一例を示す図

【図6】同実施形態に係るアンケート処理装置の動作手順の一例を示すフローチャート

【図7】同実施形態に係るアンケート処理装置の特徴抽出部の構成例を示す図

【図8】形態素つき訓練回答文集合の一例を示す図

【図9】キーフレーズベクタの一例を示す図

【図10】フリーアンケート回答文特徴ベクタ集合の一例を示す図

【図11】同実施形態に係るアンケート処理装置の特徴抽出部の動作手順の一例を示すフローチャート

【図12】同実施形態に係るアンケート処理装置の分類規則獲得部の構成例を示す図

【図13】分類規則の一例を示す図

【図14】同実施形態に係るアンケート処理装置のアンケート回答文抽出部の構成例を示す図

【図15】同実施形態に係るアンケート処理装置のアンケート回答文抽出部の動作手順の一例を示すフローチャート

【図16】同実施形態に係るアンケート処理装置のアンケート回答文抽出部のアンケート分析インタフェースのイメージ例を示す図

【符号の説明】

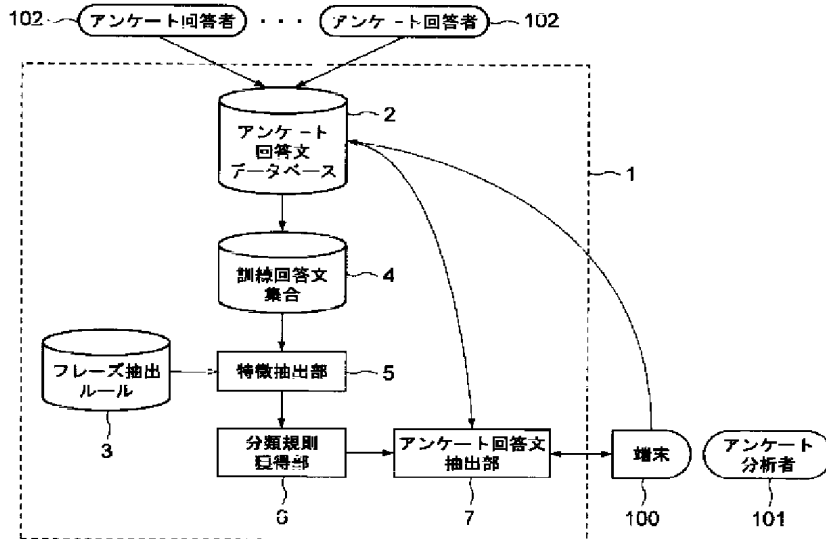
- 1…アンケート処理装置
- 2…アンケート回答文データベース
- 3…フレーズ抽出ルール
- 4…訓練回答文集合
- 5…特徴抽出部
- 6…分類規則獲得部
- 7…アンケート回答文抽出部
- 11…形態素解析部
- 12…形態素つきフリーアンケート回答文集合
- 13…キーフレーズ抽出部
- 14…キーフレーズベクタ
- 15…フォーマット変換部
- 16…フリーアンケート回答文特徴ベクタ集合
- 21…帰納学習部
- 22…分類規則データベース
- 31…推論エンジン
- 32…分析インタフェース
- 33…データベース
- 100…端末
- 201…質問文選択ボックス

202…分類カテゴリ選択ボックス
 203…検索ボタン
 204…抽出回答文表示ウインドウ

205…修正ボタン
 206…保存ボタン
 321…G U I 画面

【図1】

【図2】



Q1. 利用しているパソコンのメーカーは?
 Q2. 利用しているOSは?
 Q3. A社製品に関してご自由にお書きください
 Q4. B社製品に関してご自由にお書きください
 ...

【図10】

ID	アンケートID	特徴ベクタ	分類カテゴリ
#F1	Q3#1	11000000	苦情
#F2	Q3#2	10110000	感想+
#F3	Q3#2	01000000	苦情
#F4	Q4#1	00101100	感想+
#F5	Q4#1	00000011	感想-

【図3】

Q1	Q2	Q3	Q4	...
A社	OS_X	製品Oはよく壊れて困る。	製品Nは性能はよい。が、デザイン悪い。	...
B社	OS_X	製品Oはデザインよい。けど、すぐ壊れ困る。	製品Mは性能ピカイチ。	...
A社	OS_Y	全体的にかなりいい。	全体的にシンプル。	...
A社	OS_X	壊れて困る。	シンプルすぎるのが嫌い。	...
...

【図4】

【図5】

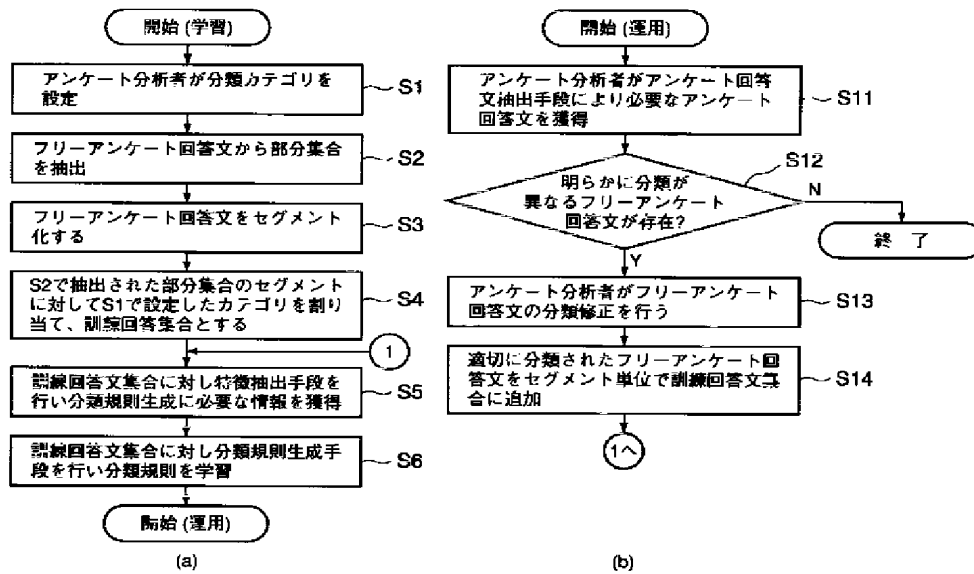
ID	フレーズ抽出ルール
#R1	{(名)}{(両名)}{(個體)}{(人面)}{(姓面)}{(地名)}{(名国)}{(の)}{(英)}
#R2	{(五)}{(下)}{(上)}{(付)}{(活尾)}
...	...

ID	アンケートID	フリーアンケート回答文	分類カテゴリ
#T1	Q3#1	製品Oはよく壊れて困る。	苦情
#T2	Q3#2	製品Oはデザインよい。	感想+
#T3	Q3#2	けど、すぐ壊れ困る。	苦情
#T4	Q4#1	製品Nは性能がよい。	感想+
#T5	Q4#1	が、デザイン悪い。	感想-

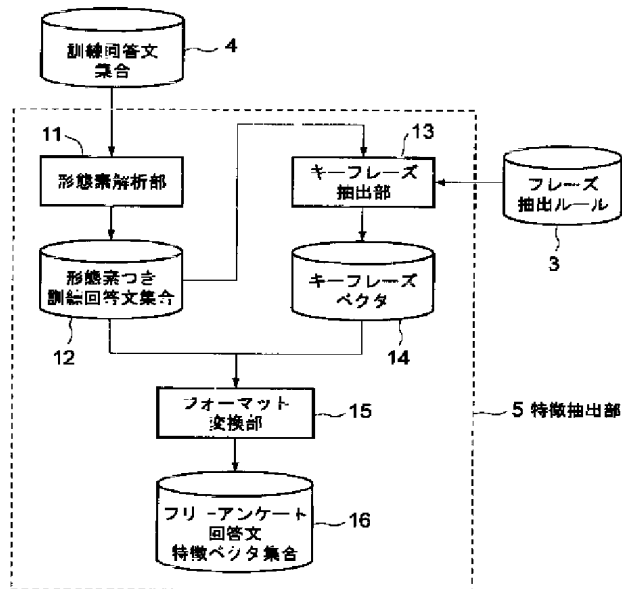
【図8】

ID	アンケートID	形態素つきフリーアンケート回答文	分類カテゴリ
#T1	Q3#1	/製品<名>/O<英>+は<付>/よ<形>+<活尾>/壊れ<下>+て<付>/困<五ら>+る<活尾>/。<句読>	苦情
#T2	Q3#2	/製品<名>/O<英>+は<付>/デザイン<両名>/よ<形>+い<活尾>/。<句読>	感想+
#T3	Q3#2	/けど<接>/。<句読>/すぐ<副>/壊れ<下>+困<五ら>/る<活尾>/。<句読>	苦情
#T4	Q4#1	/製品<名>/N<英>+は<付>/性能<名>+が<付>/よ<形>+い<活尾>/。<句読>	感想+
#T5	Q4#1	/が<付>/。<句読>/デザイン<両名>/悪<形>+い<活尾>/。<句読>	感想-

【図6】



【図7】



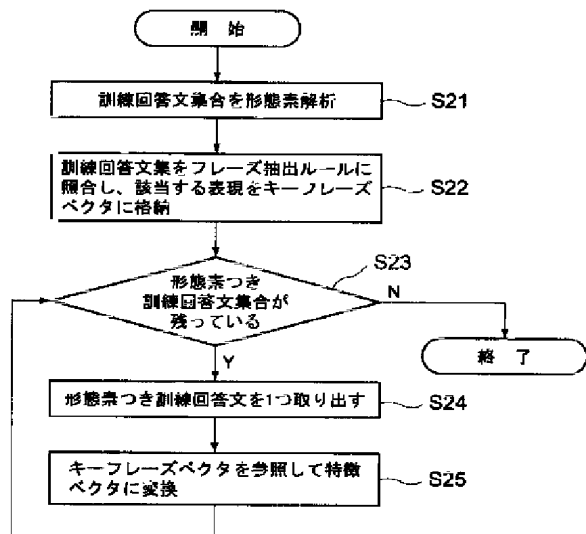
【図13】

/因<五ら>+る<活尾>	
: 1	-> 苦情
: 0	
/よ<形>+い<活尾>	
: 1	-> 迷惑+
: 0	-> 迷惑-

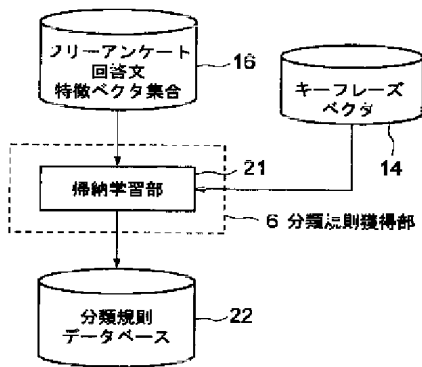
【図9】

ID	キーフレーズ	出現数
#K1	/製品<名>/O<英>	2
#K2	/因<五ら>+る<活尾>	2
#K3	/よ<形>+い<活尾>	2
#K4	/デザイン<両名>/よ<形>+い<活尾>	1
#K5	/製品<名>/N<英>	1
#K6	/性能<名>+が<付>/よ<形>+い<活尾>	1
#K7	/デザイン<両名>/悪<形>+い<活尾>	1
#K8	/悪<形>+い<活尾>	1

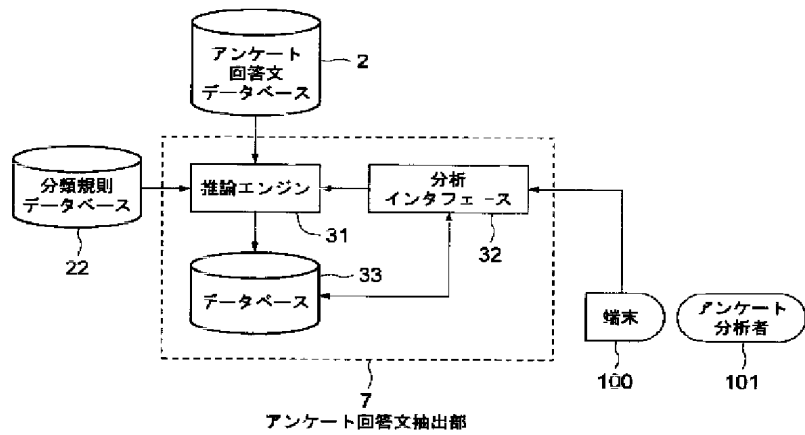
【図11】



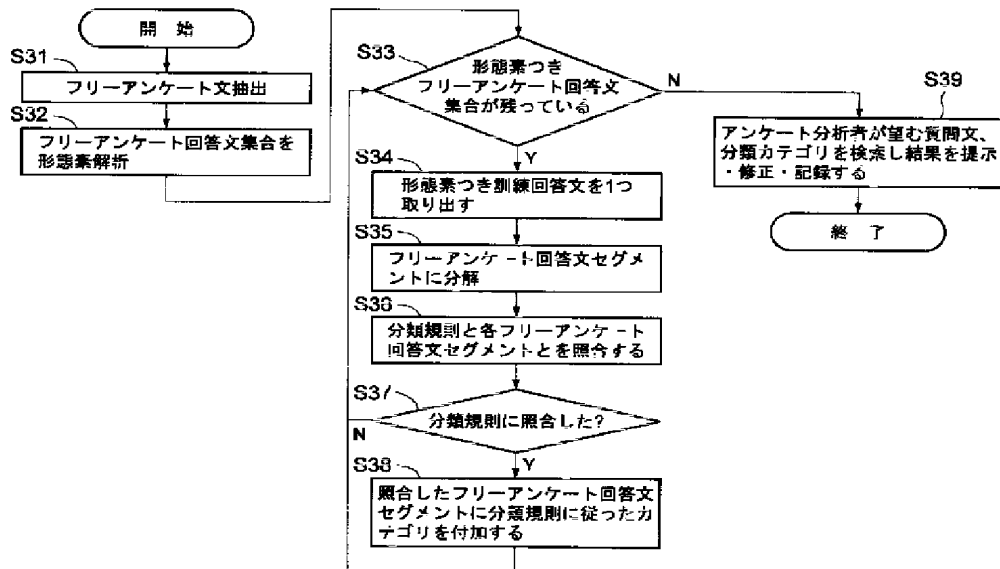
【図12】



【図14】



【図15】



【図 16】

質問番号 分類カテゴリ

Q3 苦情 検索

201 202 203

アンケート回答文

ID	フリーアンケート回答文
#A1	製品Oはよく壊れて困る。
#A2	製品Oはデザインよい。けど、すぐ壊れ困る。
...	...

204

修正 保存

205 206

321
GUI画面

フロントページの続き

(72)発明者 市村 由美
神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

(72)発明者 折原 良平
神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内
Fターム(参考) 5B075 ND03 ND20 NK06 NK31 NR12
PQ46